情報量規準による区分的線形推定

上 坂 吉 則*

Piecewise Linear Estimation by means of an Information Criterion

by

Yoshinori UESAKA

Abstract

It is well known that the parameter estimation in a linear regressive model can be easily achieved by means of the method of least-squares. There is, however, no way straightforwards connecting the method of least-squares with the parameter estimation appeared in a "piecewise" linear regressive model. The present paper gives a solution of this estimation problem by means of AIC (An Information Criterion).

1. まえがき

統計解析はいくつかのデータから母数を推定することを目的としている。その際，データの構造としていわゆる線形回帰モデル[1]がしばしば用いられる。これはいくつかのデータ $y_1, \cdots, y_n$ が未知母数 $\alpha_1, \cdots, \alpha_p$ の一次結合にノイズが加わって得られたものであるというモデルを考えるもので，式で表わせば

$$\left.\begin{aligned} y_1 &= x_{11}\alpha_1+\cdots+x_{1p}\alpha_p+\varepsilon_1, \\ &\cdots \quad \cdots \quad \cdots \quad \cdots \\ y_n &= x_{n1}\alpha_1+\cdots+x_{np}\alpha_p+\varepsilon_n \end{aligned}\right\} \tag{1}$$

となる。ここに，$\varepsilon_1, \cdots, \varepsilon_n$ は独立な確率変数（以後簡単のため rv と略記する）で平均値が 0，分散が σ^2 の正規分布に従うものである。また，x_{ij} は既知の定数である。このとき，与えられたデータ $y_1, \cdots, y_n$ から，未知母数 $\alpha_1, \cdots, \alpha_p$ を推定せよという問題が線形回帰モデルによる母数推定の問題といわれるもので，実験データからの実験式の決定などに広く応用されている。

この推定問題は母数の個数 p が既知の場合には，周知のように最小 2 乗法によって容易に解くことができる[1]。しかしながら，p が未知であったり，さらには母数がデータによって異なるような場合には最小 2 乗法で解くことは難しい。すなわち，データが区分的に未知母数の一次結合で表わされるモデル：

* 基礎工学教室

$$\left.\begin{array}{l} y(1) = x(1,1)\alpha(1,1)+\cdots+x(1,p_1)\alpha(1,p_1)+\varepsilon(1), \\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots \\ y(n_1)=x(n_1,1)\alpha(1,1)+\cdots+x(n_1,p_1)\alpha(1,p_1)+\varepsilon(n_1), \end{array}\right\} \quad (2-1)$$

$$\left.\begin{array}{l} y(n_1+1) = x(n_1+1,\ 1)\alpha(2,\ 1)+\cdots+x(n_1+1,\ p_2)\alpha(2,p_2)+\varepsilon(n_1+1), \\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots \\ y(n_2) = x(n_2,\ 1)\alpha(2,\ 1)+\cdots+x(n_2,\ p_2)\alpha(2,\ p_2)+\varepsilon(n_2), \end{array}\right\} \quad (2-2)$$

$$\begin{array}{l} \cdots\ \cdots\ \cdots\ \cdots\ \cdots \\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\cdots \end{array}$$

$$\left.\begin{array}{l} y(n_{K-1}+1) = x(n_{K-1}+1,\ 1)\alpha(K,\ 1)+\cdots+x(n_{K-1}+1,\ p_K)\alpha(K,\ p_K)+\varepsilon(n_{K-1}+1), \\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots \\ y(n_K) = x(n_K,\ 1)\alpha(K,\ 1)+\cdots+x(n_K,\ p_K)\alpha(K,\ p_K)+\varepsilon(n_K) \end{array}\right\} \quad (2-K)$$

における推定問題には最小2乗法を単純に適用することはできない。ここに，n_K 個のデータ $y(1), \cdots, y(n_K)$ は K 個の区分 $y(1), \cdots, y(n_1)$；$y(n_1+1), \cdots, y(n_2)$；$\cdots$；$\cdots$；$y(n_{K-1}+1), \cdots, y(n_K)$ に分けられ，各区分で未知母数 $\alpha(k, 1), \cdots, \alpha(k, p_k)$の一次結合にノイズ$\varepsilon(i)$ が加えられてデータが得られたというモデルになっている。上式で $x(i, j)$ は既知の定数であるが，区分の分点 $n_1, \cdots, n_{K-1}$ および各区分における母数 $\alpha(k, j)$ の個数 p_k も未知である。

このような〝区分的線形回帰モデル〟において，推定問題：

「与えられたデータ $y(1), \cdots, y(n_K)$ より

(1°) 母数 $\alpha(k, j)$, $k=1, \cdots, K$；$j=1, \cdots, p_k$,

(2°) 分点$n_1, \cdots, n_{K-1}$,

(3°) 母数の個数 p_k, $k=1, \cdots, K$,

.(4°) 区分の個数 K

を推定せよ。ただし，$\varepsilon(1), \cdots, \varepsilon(n_K)$ は平均値0，分散 σ^2 の正規分布にしたがう独立な rv で，分散 σ^2 は未知である」

が生ずる。本論文は，近年最尤推定法の一般化として注目されている情報量規準[2)]の考え方を用いてこの推定問題に対する一つの解を与えたものである。

2. 情報量規準による母数推定の考え方

この章では，情報量規準による母数推定の考え方[2),3)]をわれわれの問題に適合する形に再定式化する。そのために前章で述べた区分的線形回帰モデルの推定問題を一般化することからはじめよう。

いま，n個の rv：$Y_1, \cdots, Y_n$ が独立で，それぞれ確率密度関数（以後簡単のためpdf と略記する）：

$$g_1, \ \cdots \ , g_n$$

に従っているとする。ただし，g_i $(i=1, \cdots, n)$ は未知であるとする。さらにパラメータ θ を持つn 個の pdf の順序対：

$$f(\cdot \mid \theta) = (f_1(\cdot \mid \theta), \cdots, f_n(\cdot \mid \theta)) \quad (3)$$

のある集合を可算個用意する。ここに $f_i(\cdot \mid \theta)$ などは関数記号を表わし，実数xに対するその値を

$f_i(x \mid \theta)$ で表わす意味で，$f_i(\cdot \mid \theta)$ の・に変数が入ることを意味する。

$$\left.\begin{array}{l} F_1 = \{f(\cdot \mid \theta) \mid \theta \in \Theta_1\}\ , \\ F_2 = \{f(\cdot \mid \theta) \mid \theta \in \Theta_2\}\ , \\ \quad \cdots \quad \cdots \quad \cdots \quad \cdots \end{array}\right\} \tag{4}$$

と表わすことにする。$\boldsymbol{R}$ で実数の集合を表わし，Θ_1, Θ_2, … はそれぞれ $\boldsymbol{R}^{N1}$, $\boldsymbol{R}^{N2}$, … のある開集合で，共通部分を持っていてもよいとしておく。また，F_i に属する $f(\cdot \mid \theta)$ の各成分の pdf は θ による連続微分可能性など以下の議論で必要となる性質――これを一般に〝正則条件〟と呼ぶことにしよう――を持っていると仮定しておく。

さて上の F_i を $g=(g_1, \cdots, g_n)$ の〝モデル系〟，F_i の元を g の〝モデル〟と呼ぶことにする。そうすると，母数をデータより推定するということは，rv $Y_1, \cdots, Y_n$ を支配している pdf $g_1, \cdots, g_n$ としてモデル系 F_1, F_2, … を想定し，$Y_1, \cdots, Y_n$ の実現値(実際に与えられるデータ)$y_1=Y_1(\omega), \cdots, y_n=Y_n(\omega)$ を手掛りにして $g=(g_1, \cdots, g_n)$ に何らかの意味で最も近いモデル $f(\cdot \mid \hat{\theta})=(f_1(\cdot \mid \hat{\theta}), \cdots, f_n(\cdot \mid \hat{\theta}))$ を見いだし，そのときのモデルを規定しているパラメータ $\hat{\theta}$ を母数の推定値とするという形に表現することができる。ここに，ω はここで暗暗裡に想定している確率空間 $(\Omega, \mathfrak{A}, P)$ の Ω の元であって，$Y_1, \cdots, Y_n$ の一組のデータを得るための観測あるいは実験の指標である。

所で真の pdf : $g_1, \cdots, g_n$ とそれに対して想定されたモデルの pdf : $f_1(\cdot \mid \theta), \cdots, f_n(\cdot \mid \theta)$ の近さは情報量規準の立場では Kullback の導入した統計的情報量[5]：

$$\begin{aligned} I(f;g) &= \frac{1}{n}\sum_{i=1}^{n}\int g_i(x_i)\log\frac{g_i(x_i)}{f_i(x_i \mid \theta)}\,\mathrm{d}x_i \\ &= \frac{1}{n}\sum_{i=1}^{n}\int g_i(x_i)\log g_i(x_i)\mathrm{d}x_i - \frac{1}{n}\sum_{i=1}^{n}\int g_i(x_i)\log f_i(x_i \mid \theta)\mathrm{d}x_i \end{aligned} \tag{5}$$

で測る。この $I(f;g)$ は2組の pdf の近さを表わす指標であって，常に非負で，$I(f;g)=0$ となるのは $f=g$ のときであることが知られている[5]。したがって，$I(f;g)$ が最小となる f を g の推定としようと主張するのが情報量規準の基本的な推定理念である。式(5)で右辺の第1項は f の撰択に依存しない定数であるから，量：

$$J(f;g) = \frac{1}{n}\sum_{i=1}^{n}\int g_i(x_i)\log f_i(x_i \mid \theta)\mathrm{d}x_i \tag{6}$$

を最大ならしめる f を g の推定とすることと等価である。

以上の考察の下に，情報量規準による推定方法はつぎのように要約できる。まず，各モデル系 $F_j(j=1, 2, \cdots)$ の中から，最尤推定法によって，g に対する推定としてモデル $f^{(j)}(\cdot \mid \hat{\theta}_j)$を求める。すなわち，$f^{(j)}_1(\cdot \mid \theta)=(f^{(j)}_1(\cdot \mid \theta), \cdots, f^{(j)}_n(\cdot \mid \theta))$ とおいて，

$$\frac{1}{n}\sum_{i=1}^{n}\log f^{(j)}_i(y_i \mid \hat{\theta}_j) = \max_{\theta\in\Theta_j}\frac{1}{n}\sum_{i=1}^{n}\log f^{(j)}_i(y_i \mid \theta) \tag{7}$$

を満たす $\hat{\theta}_j$ を求める。ここに，$y_1, \cdots, y_n$ はrv $Y_1, \cdots, Y_n$ の実現値（データ）であり，Θ_j はモデル系 F_j のpdf がとり得るパラメータの領域である(式(4)参照)。この$f^{(j)}(\cdot \mid \hat{\theta}_j)$ を g に対する〝モデル系 F_j における最良モデル〟と呼ぶことにする。この段階で各モデル系における最良モデルの集合

$$\hat{F} = \{f^{(j)}(\cdot \mid \hat{\theta}_j) \mid j=1, 2, \cdots\}$$

が得られるが，この中から，最終的に g に対して最もよいモデルを撰択しなければならない。それには先の議論での式(6)を用いて，

$$J(\hat{f}; g) = \max_j J(f^{(j)}; g), \tag{8}$$

すなわち，

$$\frac{1}{n}\sum_{i=1}^{n}\int g_i(x_i)\log \hat{f}_i(x_i \mid \hat{\theta})\mathrm{d}x_i = \max_j \frac{1}{n}\sum_{i=1}^{n}\int g_i(x_i)\log f_i^{(j)}(x_i \mid \hat{\theta}_j)\mathrm{d}x_i \tag{9}$$

を満たす $\hat{f}(\cdot \mid \hat{\theta})$ を求める。このときの $\hat{\theta}$ を推定値とする立場が情報量規準での母数推定の考え方である。

上に述べた手法において式(7)の $\hat{\theta}_j$ は $f^{(j)}(\cdot \mid \theta)\epsilon F_j$ に対する適当な正則条件の下に与えられたデータ $y_1, \cdots, y_n$ から容易に求めることができる。しかしながら，式(9)によって $\hat{\theta}$ を求めるには未知の pdf である g が判明しなければならないので，実際には式(9)そのものを計算することはできない。そこで，データ $y_1, \cdots, y_n$ から式(6)の期待値を推定し，その値によって $\hat{\theta}$ を求めるという方針をとる。この期待値に関してつぎの補助定理が知られている[3]。

まず，g に対するモデル系：

$$F = \{f(\cdot \mid \theta) \mid f(\cdot \mid \theta) = (f_1(\cdot \mid \theta), \cdots, f_n(\cdot \mid \theta)),\ \theta\epsilon\Theta\} \tag{10}$$

を一つ定める。ここに，Θ は p 次元ユークリッド空間 $\boldsymbol{R}^p$ のある開集合，$f_i(\cdot \mid \theta)$ などは Θ 上で θ に関して以下の議論にとって必要な回数連続微分可能であるなど適当な正則条件を満しているものとする。つぎに，$h: \boldsymbol{R}^p \to \boldsymbol{R}$ を p 変数の正則条件を満たす関数とし，h の θ_q による偏導関数の $\theta=\tilde{\theta}$ における値を

$$\frac{\partial h}{\partial \theta_q}(\tilde{\theta}) \tag{11}$$

で表わす。同様に，h を θ_q と θ_r で2回偏微分した偏導関数の $\theta=\tilde{\theta}$ における値を

$$\frac{\partial^2 h}{\partial \theta_q \partial \theta_r}(\tilde{\theta}) \qquad (q, r=1, \cdots, p) \tag{12}$$

で表わす。さらに，これらの記法を用いて，行列と微分演算の記法をつぎのように定める。すなわち，

$$\frac{\partial^2 h}{\partial \theta \partial \theta^t}(\tilde{\theta}) = \begin{pmatrix} \frac{\partial^2 h}{\partial \theta_1 \partial \theta_1}(\tilde{\theta}) \cdots\cdots \frac{\partial^2 h}{\partial \theta_1 \partial \theta_p}(\tilde{\theta}) \\ \vdots \\ \frac{\partial^2 h}{\partial \theta_p \partial \theta_1}(\tilde{\theta}) \cdots\cdots \frac{\partial^2 h}{\partial \theta_p \partial \theta_p}(\tilde{\theta}) \end{pmatrix} \tag{13}$$

$$\frac{\partial h}{\partial \theta}(\tilde{\theta})\frac{\partial h}{\partial \theta^t}(\tilde{\theta})=\begin{pmatrix} \frac{\partial h}{\partial \theta_1}(\tilde{\theta})\frac{\partial h}{\partial \theta_1}(\tilde{\theta}) \cdots \frac{\partial h}{\partial \theta_1}(\tilde{\theta})\frac{\partial h}{\partial \theta_p}(\tilde{\theta}) \\ \vdots \\ \frac{\partial h}{\partial \theta_p}(\tilde{\theta})\frac{\partial h}{\partial \theta_1}(\tilde{\theta}) \cdots \frac{\partial h}{\partial \theta_p}(\tilde{\theta})\frac{\partial h}{\partial \theta_p}(\tilde{\theta}) \end{pmatrix} \tag{14}$$

と定める。さて，われわれの議論で重要な役割を果す，2つの $p \times p$ の行列：

$$J(\tilde{\theta})=-\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(\frac{\partial^2 \log f_i(Y_i \mid \theta)}{\partial \theta\, \partial \theta^t}(\tilde{\theta})\right) \tag{15}$$

および

$$I(\tilde{\theta})=\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(\frac{\partial \log f_i(Y_i \mid \theta)}{\partial \theta}(\tilde{\theta})\frac{\partial \log f_i(Y_i \mid \theta)}{\partial \theta^t}(\tilde{\theta})\right) \tag{16}$$

を用意する。ここに，Eは Y_i に関して期待値をとることを表わす。

さて，以上の記法のもとに補助定理はつぎのように述べられる。

〔補助定理〕与えられたデータ $y_1, \cdots, y_n$（すなわち，$Y_1, \cdots, Y_n$ の実現値）に対して

$$\frac{1}{n}\sum_{i=1}^{n}\log f_i(y_i \mid \hat{\theta})=\max_{\theta \in \Theta}\frac{1}{n}\sum_{i=1}^{n}\log f_i(y_i \mid \theta) \tag{17}$$

を満たす $\hat{\theta}$ は$y_1, \cdots, y_n$ の関数であるが，この $y_1, \cdots, y_n$ にそれぞれ $Y_1, \cdots, Y_n$ を代入したものも同じ記号 $\hat{\theta}$ で表わす。このとき，

$$\frac{1}{n}\sum_{i=1}^{n}\int g_i(x_i)\log f_i(x_i \mid \hat{\theta})\mathrm{d}x_i \tag{18}$$

の期待値*)は近似的に

$$\mathrm{AIC}'=\frac{1}{n}\sum_{i=1}^{n}\log f_i(y_i \mid \hat{\theta})-\frac{1}{n}\operatorname{trace} J(\hat{\theta})^{-1}I(\hat{\theta}) \tag{19}$$

で与えられる。ここに，trace は行列のトレースを表わす。この量 AIC′ を "情報量規準 An Information Criterion" という。

3. 区分的線形回帰モデルの推定

この章では，前章で概説した情報量規準による推定手法をわれわれの推定問題に具体的に応用してその解を与える。そのために，モデル系を設定することからはじめよう。

式(4)の形にモデル系を構成する際に pdf $f(\cdot \mid \theta)$ のパラメータ θ はそれによって $f(\cdot \mid \theta)$ が微分可能であるなどの正則条件を考慮しなければならない。したがって，第1章で述べた区分的線形回帰モデルの推定問題で述べたパラメータの内，分点，母数の個数および区分の個数は離散的であるから，θ

*) 式(18)の $\hat{\theta}$ は$\mathrm{Y}_1, \cdots, \mathrm{Y}_n$ を経由して rv であるから，式(18)も rv となる。

の成分にはなり得ない。すなわち，θとして母数と標準偏差を成分とするベクトル：

$$\theta=(\theta(1,1),\cdots,\theta(1,p_1),\cdots,\theta(K,1),\cdots,\theta(K,p_K),\sigma)^t\in \boldsymbol{R}^{\sum_{k=1}^{K}p_k+1} \tag{20}$$

を用いる*)。ここにtは行列の転置を表わす。第1章で仮定したように，式(2)で$\varepsilon(i)$は平均値0，分散σ^2に従うから，式(2)の左辺$y(i)$に対応するrv $Y(i)$は平均値が$x(i,1)\alpha(k,1)+\cdots+x(i,p_k)\alpha(k,p_k)$，分散が$\sigma^2$の正規分布に従う。このことを留意して，パラメータ：

$$\left.\begin{array}{l} K=1,2,\cdots, \\ p_1,p_2,\cdots,p_{K}=1,2,\cdots; \\ 0<n_1<n_2<\cdots<n_{K-1}<n_K\equiv n \end{array}\right\} \tag{21}$$

を持つつぎのような可算個のモデル系$F(K,p_1,\cdots,p_K,n_1,\cdots,n_{k-1})$を設定する。すなわち，

$$F(K,p_1,\cdots,p_K,n_1,\cdots,n_{K-1})$$
$$=\{f(\cdot\mid\theta)\mid f(\cdot\mid\theta)=(f_1(\cdot\mid\theta),\cdots,f_n(\cdot\mid\theta))\text{ で，}$$

$$f_i(y\mid\theta)=\frac{1}{\sqrt{2\pi\sigma}}\exp\left[-\frac{1}{2\sigma^2}(y-\mu(i))^2\right]\ (i=1,\cdots,n),$$

但し，$n_{k-1}<i\leqq n_k$なるiに対して

$$\mu(i)=\sum_{p=1}^{p_k}x(i,p)\,\theta(k,p)$$

であり，さらに$\theta(k,p)\in\boldsymbol{R}\ (k=1,\cdots,K;\ p=1,\cdots,p_k)$，$\sigma\in\boldsymbol{R}\}$ (22)

とおく。このとき，モデル系$F(K,p_1,\cdots,p_K,n_1,\cdots,n_{K-1})$の情報量規準AIC′は，後に示すように，つぎのように与えられる。

〔主定理〕式(22)のモデル系の情報量規準AIC′は

$$\begin{aligned}\mathrm{AIC}&\equiv-2n\cdot\mathrm{AIC}'-n\log 2\pi e+1\\&=n\log\hat{\sigma}^2+2\sum_{k=1}^{K}p_k+b^4/\hat{\sigma}^4\end{aligned} \tag{23}$$

で与えられる。ここに，$\hat{\sigma}^2$とb^4は

$$\hat{\sigma}^2=\frac{1}{n}\sum_{k=1}^{K}\sum_{n_{k-1}<i\leq n_k}\Big(y_i-\sum_{p=1}^{p_k}x(i,p)\hat{\theta}(k,p)\Big)^2, \tag{24}$$

$$b^4=\frac{1}{n}\sum_{k=1}^{K}\sum_{n_{k-1}<i\leq n_k}\Big(y_i-\sum_{p=1}^{p_k}x(i,p)\hat{\theta}(k,p)\Big)^4 \tag{25}$$

であり，さらに$\hat{\theta}(k,p)$らはつぎの連立一次方程式の解である。すなわち，$k=1,\cdots,K$に対して

*) ここで，式(2)の母数$\alpha(k,j)$の変数として$\theta(k,j)$なる記法を用いている。

$$\begin{pmatrix} [x(\cdot,1)x(\cdot,1)]_k \cdots [x(\cdot,1)x(\cdot,p_k)]_k \\ \vdots \\ [x(\cdot,p_k)x(\cdot,1)]_k \cdots [x(\cdot,p_k)x(\cdot,p_k)]_k \end{pmatrix} \begin{pmatrix} \hat\theta(k,1) \\ \vdots \\ \hat\theta(k,p_k) \end{pmatrix} = \begin{pmatrix} [x(\cdot,1)y(\cdot)]_k \\ \vdots \\ [x(\cdot,p_k)y(\cdot)]_k \end{pmatrix} \tag{26}$$

である。ここに，$k=1, \cdots, K$ および $1 \leqq p, q \leqq p_k$ に対して

$$\left.\begin{aligned} [x(\cdot,p)x(\cdot,q)]_k &= \sum_{n_{k-1}<i\leqq n_k} x(i,p)x(i,q), \\ [x(\cdot,p)y(\cdot)]_k &= \sum_{n_{k-1}<i\leqq n_k} x(i,p)y(i) \end{aligned}\right\} \tag{27}$$

なる記法を用いている。

この結果を用いると，第1章で述べた区分的線形回帰モデルにおける推定問題はつぎの手順で解くことができる：

(1°) モデル系を規定するパラメータ$K, p_1, \cdots, p_K, n_1, \cdots, n_{K-1}$ の1組を設定する。

(2°) 与えられたデータより，式(26)を解いて $\hat\theta(k,1), \cdots, \hat\theta(k,p_k)$ を求める。但し，$k=1, \cdots, K$ である。

(3°) この結果を式(24), (25) に代入して $\hat\sigma^2$ および b^4 を求める。

(4°) この結果を式(23)に代入して AIC を求める。

以上の手順でわかるように，AIC はパラメータ$K, p_1, \cdots, p_K, n_1, \cdots, n_{K-1}$ の関数となるので，これを $\mathrm{AIC}(K, p_1, \cdots, p_K, n_1, \cdots, n_{K-1})$ と書くことにしよう。さて，データの測定状況などから，$K, p_1, \cdots, p_K$ の動き得る範囲を十分大きくとり，上の (1°) — (4°) で得たAIC のうちの最小のものを求める。すなわち，

$$\begin{aligned} \mathrm{AIC}(K^*, p_1^*, \cdots, p_K^*, \cdots, n_1^*, \cdots, n_{K-1}^*) = \min\{&\mathrm{AIC}(K, p_1, \cdots, p_K, n_1, \cdots, n_{K-1}) \mid 1 \leqq K \leqq \widetilde{K}, \\ &1 \leqq p_1 \leqq \widetilde{p}_1, \cdots, 1 \leqq \widetilde{p}_K \leqq p_K, 0 < n_1 < \cdots < n_{K-1} < n_K \equiv n\} \end{aligned} \tag{28}$$

を満たす $K,^* p_1^*, \cdots, p_K^*, n_1^*, \cdots, n_{K-1}^*$ を求める。ここに，$\widetilde{K}, \widetilde{p}_1, \cdots, \widetilde{p}_K$ はデータの測定状況などから設定する $K, p_1, \cdots, p_K$ の上界である。このとき，モデル系 $F(K,^* p_1^*, \cdots, p_K^*, n_1^*, \cdots, n_{K-1}^*)$ が情報量規準の意味で最良のモデルとして採用され，そのモデル系における式(26)の解が求めるべき母数の推定値である。

この章の残りの部分で，主定理の式(23)の導出を順を追って述べることにする。

〔1〕式(24)と(26)の導出。まず，式(22)に注意して，$n_{k-1} < i \leqq n_k$ なる i に対して，

$$\log f_i(y(i) \mid \theta) = -\frac{1}{2\sigma^2}\Bigl(y(i) - \sum_{p=1}^{p_k} x(i,p)\theta(k,p)\Bigr)^2 - \frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 \tag{29}$$

となるから，

$$\begin{aligned} L(\theta) &\equiv \frac{1}{n}\sum_{i=1}^{n} \log f_i(y(i) \mid \theta) \\ &= -\frac{1}{2n\sigma^2}\sum_{k=1}^{K}\sum_{n_{k-1}<i\leqq n_k}\Bigl(y(i) - \sum_{p=1}^{p_k} x(i,p)\theta(k,p)\Bigr)^2 - \frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 \end{aligned} \tag{30}$$

を得る。ここで，θ は式(20)の形をしているから，式(17)にしたがって，上の $L(\theta)$を最大ならしめる θ

は $\partial L/\partial\theta=0$ を満たす。ところが,

$$\frac{\partial L}{\partial\theta(k,p)}=\frac{1}{n\sigma^2}\sum_{n_{k-1}<i<n_k}\{x(i,\ p)\ (y(i)-\sum_{r=1}^{p_k}x(i,\ r)\theta(k,\ r)\ )\}$$
$$=\frac{1}{n\sigma^2}\ \{[x(\cdot,\ \ p)p(\cdot)]_k-\ \sum_{r=1}^{p_k}\ [x(\cdot,\ \ p)x(\cdot,\ \ r)]_k\theta(k,\ \ r)\}\ \begin{pmatrix}k=1,\ \cdots,\ K\,;\\ p=1,\ \cdots,\ p_k\end{pmatrix} \tag{31}$$

および

$$\frac{\partial L}{\partial\sigma}=\frac{1}{n\sigma^3}\sum_{k=1}^{K}\sum_{n_{k-1}<i\leqq n_k}(y(i)-\sum_{p=1}^{p_k}x(i,\ p)\theta(k,p)\ )^2-\frac{1}{\sigma} \tag{32}$$

であるから, 式(31)を 0 とおいて式(26)を得, またその解を式(32)に代入し, σ^2 について解くことにより, 式(24)を得る。

〔2〕式(15)に相当する $J(\hat{\theta})$ の計算。まず,

$$J(\tilde{\theta})=\frac{1}{n}\sum_{k=1}^{K}\sum_{n_{k-1}<i\leqq n_k}\mathrm{E}\left(\frac{\partial^2\log f_i(Y(i)\mid\theta)}{\partial\theta\partial\theta^t}(\tilde{\theta})\right) \tag{33}$$

と書けるが, 微分を忠実に実行すると

$$\sum_{k=1}^{K}\sum_{n_{k-1}<i\leqq n_k}\left(\frac{\partial^2\log f_i(Y(i)\mid\theta)}{\partial\theta\partial\theta^t}\ (\tilde{\theta})\right)=-\frac{1}{\tilde{\sigma}^2}\begin{pmatrix}A_1 & & 0 & b_1\\ & \ddots & & \vdots\\ & & \ddots & \vdots\\ 0 & & A_K & b_K\\ b_1^t & \cdots & b_K^t & c\end{pmatrix} \tag{34}$$

を得る。ここに, A_k は$p_k\times p_k$ の行列で式(26)の左辺に現われる$[x(\cdot,\ p)x(\cdot,\ q)]_k$ を要素とする行列であり, b_k と c はそれぞれ

$$b_k=\frac{2}{\tilde{\sigma}}\ \Big(\sum_{n_{k-1}<i\leqq n_k}(Y(i)-\sum_{p=1}^{p_k}x(i,\ p)\tilde{\theta}(k,\ p)\ )x(i,\ 1),\ \cdots,$$
$$\sum_{n_{k-1}<i\leqq n_k}(Y(i)-\sum_{p=1}^{p_k}x(i,\ p)\tilde{\theta}(k,p))x(i,\ p_k)\Big)^t \tag{35}$$

なる $p_k\times 1$ の行列と

$$c=\sum_{k=1}^{K}\sum_{n_{k-1}<i\leqq n_k}\left\{\frac{3}{\tilde{\sigma}^2}(Y(i)-\sum_{p=1}^{p_k}x(i,\ p)\tilde{\theta}(k,\ p)\ )^2-1\right\} \tag{36}$$

なる 1×1 の行列である。ここに, 式(20)に対応して

$$\tilde{\theta}=(\tilde{\theta}(1,\ 1),\ \cdots,\ \tilde{\theta}(K,\ p_K),\ \tilde{\sigma})^t \tag{37}$$

である。さて,〔1〕で求めた $\hat{\theta}$ は母数 $\alpha(k,\ p)$ $(k=1,\ \cdots,\ K\ ;\ p=1,\ \cdots,\ p_k)$ と σ の最尤推定量であるから, 近似的に $\hat{\theta}(k,\ p)\fallingdotseq\alpha(k,\ p)$, $\hat{\sigma}^2\fallingdotseq\sigma^2$ と考えられ,

$$\mathrm{E}\ \left(Y(i)-\sum_{p=1}^{p_k}x(i,\ p)\hat{\theta}(k,p)\right)\ \fallingdotseq\mathrm{E}\ \left(Y(i)-\sum_{p=1}^{p_k}x(i,\ p)\alpha(k,\ p)\ \right)=0, \tag{38}$$

$$\mathrm{E}\ \left(Y(i)-\sum_{p=1}^{p_k}x(i,\ p)\hat{\theta}(k,p)\right)^2\fallingdotseq\mathrm{E}\ \left(Y(i)-\sum_{p=1}^{p_k}x(i,\ p)\alpha(k,\ p)\ \right)^2$$

$$=\mathrm{E}(Y(i)-\mathrm{E}\,Y(i))^2=\sigma^2\fallingdotseq\tilde{\sigma}^2 \tag{39}$$

を得る。したがって，b_k, c の $\tilde{\theta}$ に $\hat{\theta}$ を代入したものの期待値は近似的に

$$\mathrm{E}b_k\fallingdotseq(0,\cdots,0)^t,\quad \mathrm{E}c=\sum_{k=1}^{K}2(n_k-n_{k-1})=2n \tag{40}$$

となる。さらに，明らかに，$\mathrm{E}A_k=A_k$ であるから，式(34)の $\tilde{\theta}$ に $\hat{\theta}$ を代入して期待値をとることによって，結局，

$$J(\hat{\theta})=\frac{1}{n\hat{\sigma}^2}\begin{pmatrix} A_1 & & & 0 \\ & \ddots & & \\ & & A_K & \\ 0 & & & 2n \end{pmatrix} \tag{41}$$

を得る。

〔3〕式(16)に相当する $I(\hat{\theta})$ の計算。まず，

$$I(\tilde{\theta})=\frac{1}{n}\sum_{k=1}^{K}\sum_{n_{k-1}<i\leq n_k}\mathrm{E}\left(\frac{\partial\log f_i(Y(i)\mid\theta)}{\partial\theta}(\tilde{\theta})\frac{\partial\log f_i(Y(i)\mid\theta)}{\partial\theta^t}(\tilde{\theta})\right) \tag{42}$$

と書けるが，微分を忠実に実行すると

$$\sum_{k=1}^{K}\sum_{n_{k-1}<i\leq n_k}\left(\frac{\partial\log f_i(Y(i)\mid\theta)}{\partial\theta}(\tilde{\theta})\frac{\partial\log f_i(Y(i)\mid\theta)}{\partial\theta^t}(\tilde{\theta})\right)=\frac{1}{\tilde{\sigma}^2}\begin{pmatrix} [d_1d_1^t]_1 & & 0 & \\ & \ddots & & \boxed{*} \\ 0 & & [d_Kd_K{}^t]_K & \\ & \boxed{*} & & e^2 \end{pmatrix} \tag{43}$$

を得る。ここに，

$$\alpha_k=\left(\frac{1}{\tilde{\sigma}}\left(Y(i)-\sum_{p=1}^{p_k}x(i,p)\tilde{\theta}(k,p)\right)x(i,1),\right.$$
$$\left.\cdots,\frac{1}{\tilde{\sigma}}\left(Y(i)-\sum_{p=1}^{p_k}x(i,p)\tilde{\theta}(k,p)\right)x(i,p_k)\right)^t,\ (n_{k-1}<i\leq n_k) \tag{44}$$

$$[d_kd_k{}^t]_k=\sum_{n_{k-1}<i\leq n_k}d_kd_k^t, \tag{45}$$

また，

$$e^2=\sum_{k=1}^{K}\sum_{n_{k-1}<i\leq n_k}\left\{\frac{1}{\tilde{\sigma}^2}\left(Y(i)-\sum_{p=1}^{p_k}x(i,p)\tilde{\theta}(k,p)\right)^2-1\right\}^2 \tag{46}$$

である。また，$\boxed{\quad *\quad}$ は非零要素を表わす。ここで，〔2〕で用いたと同様の近似を用いると

$$\mathrm{E}\left(Y(i)-\sum_{p=1}^{p_k}x(i,p)\hat{\theta}(k,p)\right)^2\fallingdotseq\hat{\sigma}^2$$

となるから，d_k と e^2 の $\tilde{\theta}$ に $\hat{\theta}$ を代入して期待値をとると

$$\mathrm{E}[d_kd_k^t]_k\fallingdotseq A_k, \tag{47}$$

$$\mathrm{E}e^2 \fallingdotseq \sum_{k=1}^{K} \sum_{n_{k-1}<i\leqq n_k} \left(\frac{1}{\hat{\sigma}^4}\mathrm{E}\Big(Y(i)-\sum_{p=1}^{p_k} x(i,\ p)\hat{\theta}(k,\ p)\Big)^4 - 1 \right) \tag{48}$$

を得る。これらを式(42)で $\tilde{\theta}$ の代りに $\hat{\theta}$ とおいたものに代入すれば，結局，

$$I(\hat{\theta}) = \frac{1}{n\hat{\sigma}^2} \left(\begin{array}{ccc|c} A_1 & & 0 & \\ & \ddots & & * \\ 0 & & A_K & \\ \hline & * & & \mathrm{E}e^2 \end{array} \right) \tag{49}$$

を得る。

〔4〕AIC の計算。まず，式(30)の左辺の θ に $\hat{\theta}$ を代入し，式(24)の $\hat{\sigma}^2$ を用いると

$$\frac{1}{n}\sum_{i=1}^{n} \log f_i(y(i) \mid \hat{\theta}) = -\frac{1}{2\hat{\sigma}^2} \times \hat{\sigma}^2 - \frac{1}{2}\log 2\pi - \frac{1}{2}\log\hat{\sigma}^2 = -\frac{1}{2}\log\hat{\sigma}^2 - \frac{1}{2}\log 2\pi e \tag{50}$$

を得る。また，式(41)と(49)より $J^{-1}(\hat{\theta})\,I(\hat{\theta})$ は

$$J^{-1}(\hat{\theta})I(\hat{\theta}) = \left(\begin{array}{ccc|c} I_1 & & 0 & \\ & \ddots & & * \\ 0 & & I_K & \\ \hline & * & & \mathrm{E}e^2/2n \end{array} \right) \text{，但し } I_k \text{ は } p_k \text{ 次元の単位行列} \tag{51}$$

となるから，

$$\mathrm{trace} J^{-1}(\hat{\theta})I(\hat{\theta}) = \sum_{k=1}^{K} p_k + \frac{1}{2n}\mathrm{E}e^2 \tag{52}$$

を得る。ここで，式(48)で $\mathrm{E}\,Y(i)$ の不偏推定値として $Y(i)$の実現値 $y(i)$ を用いれば，

$$\mathrm{trace} J^{-1}(\hat{\theta})I(\hat{\theta}) = \sum_{k=1}^{K} p_k + \frac{1}{2n\hat{\sigma}^4} \sum_{k=1}^{K} \sum_{n_{k-1}<i\leqq n_k} \Big(y(i) - \sum_{p=1}^{p_k} x(i,\ p)\hat{\theta}(k,\ p)\Big)^4 - \frac{1}{2} \tag{53}$$

となる。そこで，式(50)と(53)を補助定理の式(19)に代入し，$\mathrm{AIC} = -2n\mathrm{AIC}' - n\log 2\pi e + 1$ とおくことにより，主定理の式(23)を得るわけである。

4. おわりに

本論文では最尤推定法の一般化である情報量規準の考え方を用いて区分的線形回帰モデルの推定問題に対する具体的な解を与えた。区分の個数 K が1の場合には，主定理から明らかなように，勿論通常の最小2乗法に帰着する。しかし，その場合でも，母数の個数 p_1 が未知の場合には AIC が最小となる p_1 を求めることによって，母数の個数を決定することができる*)。

区分的線形回帰モデルは巾広い応用を持つものであるが，その1つは $i=1, \cdots, n$ を時刻と考え，時間的に変動するデータ $y_1, \cdots, y_n$ に対する実験式を求める問題がある。この場合，区間 $(n_{k-1},\ n_k]$ で y_i が p_k-1 次の多項式にしたがっているというモデルを考えると式(2)の $x(i,\ p)$ として

*) この場合はすでに赤池によって指摘されている。[6]。

$$x(i,\ p)=i^{p-1},\ (n_{k-1}<i\leqq n_k;\ p=1,\ \cdots,\ p_k) \tag{54}$$

ととればよいことになる。そして式(26)の左辺の行列の $(p,\ q)$ 要素は

$$[x(\cdot,\ p)x(\cdot,\ q)]_k=\sum_{n_{k-1}<i\leqq n_k} i^{p+q-2} \tag{55}$$

と簡単になるので，この行列の逆行列を一度求めておけば，種々のデータに対して適用できることになる。

なお，本論文で述べた推定法の数値計算上の難点は式(28)からわかるように，AIC の最小を求めるのに，$0<n_1<\cdots<n_{k-1}<nK\equiv n$ を満たす $n_1,\ \cdots,\ nK-1$ のすべての組合せについて調べる必要があることである。K が少し大きくなればその組合せ数は非常に大きくなり，いかにコンピュータを以ってしてもそのための計算時間は莫大なものとなる。しかしながら，動的計画法の手法を用いて AIC の最小値を探索するアルゴリズムを考案すると計算時間を大巾に少なくでき，実用上は支障なくなる[7]。この点については稿を改めて論じたいと思っている。

本研究は一部には文部省科学研究費補助金特別研究(1)課題番号 911305 に基づくものである。

参 考 文 献

1）森口繁一：統計解析，岩波講座現代応用数学 B. 10-b., 岩波書店，1957.

2）H. Akaike : Information theory and an extension of the maximum likelihood principle, in "B. N. Petrov and F. Csáki (eds.): 2nd International Symposium on Information Theory, Akadémiai Kiadó, Budapest, 1973", pp. 267—281.

3）竹内啓：情報統計量の分布とモデルの適切さの規準，数理科学，No. 153，3 月号，pp. 12—18, 1976.

4）R. H. Nordon : A survey of maximum likelihood estimation, Int. Stat. Rev., Part 1—Vol. 40, No. 3, pp. 329—354, 1972 ; Part 2—Vol. 41, No. 1, pp. 39—58, 1973.

5）Kullback : Information and statistics, Dover Publications, 1968.

6）赤池弘次：情報量規準 AIC とは何か，数理科学，No. 153，3 月号，pp. 5—11, 1976.

7）上坂吉則：図形分節のアルゴリズムとその応用，文部省特定研究「生体の制御情報システム」研究論文，No. 73, 1976—7—23.

（昭和51年 8 月11日受理）